

The Reanalysis for the Global Ensemble Forecast System, version 12.

Thomas M. Hamill¹, Jeffrey S. Whitaker¹, Anna Shlyueva², Gary Bates³, Sherrie Fredrick,³ Philip Pegion³, Eric Sinsky⁶, Yuejian Zhu⁴, Vijay Tallapragada⁴, Hong Guan⁵, Xiaqiong Zhou⁴, and Jack Woollen⁴

¹ *NOAA Physical Sciences Laboratory, Boulder, CO*

² *Joint Center for Satellite Data Assimilation, University Corporation for Atmospheric Research, Boulder CO*

³ *Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder CO*

⁴ *National Weather Service, Environmental Modeling Center, College Park, MD*

⁵ *SRG at NOAA/NWS/NCEP/EMC, College Park, MD*

⁶ *IMSG at NOAA/NWS/NCEP/EMC, College Park, MD*

Submitted to *Monthly Weather Review* (presumably)

9 February 2021

Corresponding author:

Dr. Thomas M. Hamill,
NOAA ESRL/PSL, R/PSL1
325 Broadway, Boulder CO 80305
tom.hamill@noaa.gov
phone: +1 (303) 497-3060
telefax: +1 (303) 497-6449
telework: +1 (303) 516-0340

24

25
26
27
28
29
30
31
32
33
34
35
36
37

38

39

ABSTRACT

NOAA has created a global reanalysis data set, intended primarily for initialization of reforecasts for its Global Ensemble Forecast System, version 12 (GEFSv12), which provides ensemble forecasts out to +35 days lead time. The reanalysis covers the period 2000-2019. It assimilates most of the observations that were assimilated into the operational data assimilation system used for initializing global predictions. These include a variety of conventional data, infrared and microwave radiances, Global Positioning System radio occultations, and more. The reanalysis quality is generally superior to that from NOAA’s previous-generation Climate Forecast System Reanalysis (CFSR), demonstrated in the fit of short-term forecasts to the observations and in the skill of 5-day deterministic forecasts initialized from CFSR vs. GEFSv12. Skills of reforecasts initialized from the new reanalyses are similar but slightly lower than skills initialized from a pre-operational version of the real-time data assimilation system conducted at the higher, operational resolution. Basic control reanalysis data are made publicly available.

1. Introduction.

The Global Ensemble Forecast System (GEFS) produced by the US National Weather Service (NWS) is one of several prediction systems conducted by the NWS. It supports medium-range weather and sub-seasonal to seasonal forecasting. The newest version 12 of the GEFS (GEFSv12 hereafter) produces 31-member ensemble forecasts each day to +35 days lead time. The GEFSv12 system configuration and performance are documented in companion articles (Zhou et al. 2021). Like other prediction systems in the NWS, this system is typically updated every several years.

As current-generation ensemble guidance is limited by systematic errors including resolution limitations, errors in the mean state, and errors in the ensemble spread (standard deviation about the mean), products from the GEFS are commonly statistically post-processed. A time series of past ensemble forecasts and coincident observations / analyses are used to estimate the systematic errors (Vannitsem et al. 2018) and adjust the real-time forecast, improving skill and reliability. Statistical postprocessing can also filter the predictable signal from the meteorological noise due to chaotic error growth and sampling variability due to finite ensemble size. When used in combination with higher-resolution analyses, statistical postprocessing of ensemble forecasts can also provide a downscaling related to the modulation of weather by physiographic features (Hamill and Whitaker 2006).

For infrequent events and longer-lead forecasts where skill is marginal, the statistical postprocessing is greatly aided from a long training data set of reforecasts, i.e., retrospective forecasts using the same prediction system used to generate the real-time forecasts (Hamill et al. 2004, 2006, 2008, 2013, Hamill and Whitaker 2006, Hagedorn et al. 2008, Scheuerer and Hamill 2015). Similarity of the statistical characteristics of the reforecasts and the real-time forecasts is highly desirable for statistical postprocessing; in

64 this way the training data resembles the real-time forecast data. When this statistical
65 consistency is lost, statistically post-processed product quality may suffer, especially the
66 postprocessing of shorter-lead forecasts (Hamill 2017). Statistical discrepancies in the
67 short-term forecasts may occur in part because of the differing characteristics of the initial
68 conditions in the reanalysis vs. the real-time analysis. For this reason, initialization of
69 reforecasts from an archive of past operational initial states from obsolete versions of
70 modeling systems is unlikely to provide statistical consistency; the initial conditions will
71 reflect in part the systematic errors of the outdated prediction system versions used to
72 provide the background forecasts in the data assimilation. Hence, if computationally
73 feasible, production of a new set of retrospective analyses (reanalyses) consistent with the
74 operational system is highly desirable for reforecast initialization.

75 Many organizations across the globe have created global, multi-decadal reanalyses.
76 These are retrospective gridded analyses of the state of the atmosphere (and sometimes
77 other state components such as the land, ocean, and sea ice), commonly produced with a
78 cycled data assimilation system leveraging a prior background forecast updated to the newly
79 available observations. Global reanalyses have been provided by the the Copernicus
80 Climate Service/European Centre for Medium-Range Weather Forecasts (ECMWF; Uppala
81 et al. 2005, Dee et al. 2011, Laloyaux et al. 2018, Hersbach et al. 2019), the Japan
82 Meteorological Agency (Kobayashi et al. 2015, Harada et al. 2016), the National
83 Aeronautics and Space Administration (NASA, Rienecker et al. 2011, Gelaro et al. 2017),
84 the National Centers for Environmental Prediction (NCEP) and the National Centers for
85 Atmospheric Research (NCAR; Kalnay et al. 1996), and the NCEP/Department of Energy
86 (Kanamitsu et al. 2002). Several have been generated by organizations within the National
87 Oceanic and Atmospheric Administration, including the Climate Forecast System Reanalysis
88 (CFSR; Saha et al. 2010) and several versions of the 20th Century Reanalysis (20CR;

89 Compo et al. 2011, Slivinski et al. 2019). Most of these reanalyses, such as
90 Copernicus/ECMWF's "ERA5", are intended primarily for weather and climate monitoring,
91 and their design and choice of observations to assimilate emphasizes this. ERA5, for
92 example, creates a near-surface temperature analysis using 2-m temperature observations
93 alongside the full 3-D atmospheric analysis. This temperature analysis has many
94 applications, but it is used only indirectly in the initialization of ECMWF's medium-range
95 forecasts (to affect the soil-state estimate).

96 The main purpose of the reanalysis discussed in this paper, the reanalysis for the
97 GEFSv12 is *reforecast initialization*. The GEFSv12 reanalysis was designed so that its
98 initial conditions will be consistent with the operational data assimilation system used to
99 initialize the real-time GEFSv12 ensemble predictions, to the extent practical. If a user
100 seeks a reanalysis for other applications such as climate monitoring or evaluation over a
101 longer period of time than the 2000-2019 period covered by GEFSv12, they may be better
102 served by other reanalyses such as ERA5 or 20CR or MERRA-2, depending on the
103 application and the length of record needed. The GEFSv12 reanalysis is an atmosphere-
104 only reanalysis, for reasons discussed later.

105 While designed to facilitate GEFSv12 reforecast initialization for statistical
106 postprocessing, we envision this reanalysis having several other applications. For
107 example, suppose in the coming years that developers will need to initialize experimental
108 forecasts for high-impact events such as hurricane Katrina or Sandy. The GEFSv12
109 reanalysis will be the initialization data set of choice for the next several years of
110 retrospective forecast experiments for the GEFS and its deterministic counterpart, the GFS
111 (Global Forecast System). The reanalysis will provide suitable initial conditions consistent
112 with the underlying prediction system, and the reforecasts will provide a forecast

performance baseline. Eventually the statistical character of the real-time initial conditions will change, be it due to prediction-system improvements which alter the background bias (Hamill 2017) or assimilation system improvements, or more observations. At this future point, another reanalysis will be necessary; with NOAA's migration to cloud compute and storage, perhaps this next reanalysis will be computed there.

As will be discussed, ensuring statistical consistency between the real-time analyses and the reanalyses was challenging. Generating a real-time analysis is itself computationally expensive when high analysis resolution and modern data assimilation techniques are used, techniques such as 4-dimensional variational (4D-Var; Courtier et al. 1994, Rabier et al. 2000) methods or 4-dimensional ensemble-variational analysis (4D-En-Var; Wang and Lei 2014, Kleist and Ide 2015). Even with a reduced-resolution version of the operational system and with splitting the reanalysis production into streams (parallel cycles processing different segments of multi-decadal period), it took an extended period of time, $O(1 \text{ year})$ and $O(1000)$ cores per stream, to generate this uncoupled reanalysis. This necessitated production of the reanalysis before the final configuration of the GEFSv12 and the operational data assimilation system were finalized, leading to some slight inconsistencies between the GEFSv12 reanalysis configuration and the real-time GEFSv12 configuration.

This article, then, describes the GEFSv12 reanalysis. Diagnostics will focus on the general quality of the reanalysis compared to its NOAA predecessor, the CFSR. It will also provide some data to compare the characteristics of the reanalysis compared to contemporaneous real-time analyses produced for the GFS and GEFS v12 forecast initialization. The paper is organized as follows. Section 2 will briefly describe the data that were assimilated to produce the reanalysis. Section 3 provides a brief description of the

assimilation system, the forecast model, the ensemble system characteristics, and the land, ocean and sea-ice initialization procedures. Section 4 describes the characteristics of the reanalysis through diagnostics such as the fit of short-term forecasts to observations and the accuracy of medium-range deterministic forecasts initialized with GEFSv12 vs the preceding CFSR. It will also provide some diagnostics of the character of initial conditions and forecasts initialized from GEFSv12 analyses vs. the operational analyses and pre-production parallel simulations of the GEFSv12, known as “retro” runs. Section 5 reviews known issues with the reanalysis, and section 6 describes the data storage before section 7 concludes.

2. Description of the data assimilated in the GEFS v12 reanalysis.

To simplify the data processing, it was decided to leverage the observational data that had already been preprocessed to generate other previous reanalyses. Before 1 January 2016, observations collected for input to the CFSR reanalysis (Saha et al. 2010) were used. After this date, observations from the operational data stream were used. The reader is referred to this publication for a review of these data.

Satellite radiance data now provides the majority of the observations assimilated and have the greatest impact on the reanalysis quality (Gelaro et al. 2010), and the GEFSv12 reanalysis used a large number of channels from a variety of satellites. These are synthesized in Table O1 of the [online appendix](#). Generally, the microwave and IR radiances were assimilated from a number of geostationary and polar-orbiting satellites from the US and Europe. For ozone observations, SBUV/2 (Solar Backscatter Ultraviolet Version 2) observations from NOAA satellites were used for ozone profile and OMI (Ozone Monitoring Instrument) Aura for total ozone observations (see online appendix Table O2 for details).

One observational data source was problematic, namely the velocity azimuth display (VAD; Browning and Wexler 1968, Lee et al. 2014) winds. While the reanalyses were being computed, we discovered that there was insufficient quality control of these data in earlier streams, and sometimes the low-level winds were contaminated by bird migrations. We did not have the computational capacity nor time to re-start the computations from the beginning of the cycles. Accordingly, when discovered, assimilation of these wind observations were turned off. For the 1999 stream, VAD data were not assimilated after 00 UTC 21 February 2001. For the 2003 stream, VAD data were not assimilated after 18 UTC 6 January 2005. For the 2007 stream, VAD data were not assimilated after 18 UTC 27 November 2008. For other streams, QC was deemed acceptable, and their assimilation was continued.

3. Data assimilation and prediction system.

Table 1 summarizes the major differences between the CFSR and the GEFSv12 reanalysis systems. The most significant changes included use of a new atmospheric dynamical core to provide background forecasts, discussed in section 3.a below. Another major difference was that the GEFSv12 reanalyses excluded the generation of accompanying 3-dimensional ocean reanalyses for weakly coupled forecast initialization. Instead, previously produced time series of optimal interpolation (OI) sea-surface temperature (SST) analyses provided the ocean boundary condition for the reanalyses. This simplification was made in part because as the GEFSv12 real-time forecasts and reforecasts extend to +35 days lead time, changes in the ocean state were smaller in magnitude during this first month than they were for the seasonal forecasts produced by the previous Climate Forecast System version 2 (CFSv2). How SSTs evolve in the GEFSv12

184 reforecasts will be discussed in the accompanying article on the GEFS forecast and
185 reforecast procedure (Zhou et al. 2021, Guan et al. 2021).

186 The reanalyses were conducted in parallel streams of five or more years in length,
187 starting in 1999, 2003, 2007, 2011, and 2015. As soil moisture is quite sensitive to its
188 initialization and takes a long period to stabilize, the first year of each stream was discarded.
189 Hence the 1999 stream provides the reanalysis initial conditions during 2000-2003, and the
190 2003 stream provides them during 2004-2008, and so forth (Rienecker et al. 2011). An
191 examination of the time-mean soil moisture for three regions with strong land-atmosphere
192 coupling (Koster et al. 2004) are shown in Fig. 1. These three regions are the US Southern
193 Great Plains, northern equatorial Africa, and India. The time- and domain mean soil
194 moistures in these three areas do not exhibit a temporal discontinuity between stream
195 boundaries (left column). Similarly, scatterplots of beginning-of-stream vs. end-of-stream
196 soil moisture at each grid point in the three domains are shown in the right-hand column,
197 with extremely high correlation and little sign of bias. From this we infer that the separation
198 of the analysis procedure into streams appears to provide a legitimate way to parallelize the
199 reanalysis productions without degrading near-surface analysis quality.

200 *a. Atmospheric dynamical core and physical parameterization suite used for background*
201 *forecasts.*

202 The FV3 dynamical core (Lin 2004, Putman and Lin 2007) is used in both the
203 deterministic Global Forecast System (GFS) and in this version 12 of the Global Ensemble
204 Forecast System (GEFS) that was used as the forward model in the cycled data
205 assimilation. Predictions using this dynamical core were much less diffusive, containing
206 more small-scale variability relative to the previous spectral dynamical core. The FV3
207 dynamical core is also capable of cloud-resolving simulations when used at higher

resolutions (Putman and Lin 2007). In the FV3-based GFS and GEFS, grid spacings are denoted with a “CZZZ” notation, denoting that there are $ZZZ \times ZZZ$ grid boxes across each of the six faces on the cube upon which the earth’s spheroid is projected. For the C384 grid spacing, this would indicate that there are 384×384 grid points around a latitude circle, or an effective grid spacing of approximately 0.23 degrees.

The parameterization suite used in the FV3 was largely the same as that used in the Global Forecast GFS version 15. Basic details on the parameterization suite, with associated references, can be found at https://dtcenter.ucar.edu/GMTB/v3.0/sci_doc/GFS_v15_page.html and https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gefs_v12.php.

b. Data assimilation procedure.

We now consider the assimilation procedure (Kleist et al. 2018). Ideally, the exact same assimilation procedure, types of observations, and prediction system would be used in the real-time and reanalysis systems. Because of the computational expense of producing multi-decadal ensemble reanalyses, it was decided that the control reanalysis would be computed at half the resolution of the operational system, C384 vs. C768. Similarly, the ensemble providing background-error covariances was computed at C128 in the reanalysis vs. C384 in the operational system. The impacts will be examined in section 4, where diagnostics of the assimilation and forecast quality are examined.

The data assimilation procedure used in the GEFSv12 reanalysis was a hybrid 4D-ensemble-variational algorithm (4D-En-Var, Kleist and Ide 2015). This was used to update the ensemble mean, while an Local Ensemble Transform Kalman Filter (LETKF) variant of the EnKF was used to update the ensemble perturbations (Lei et al. 2018). The GEFSv12

reanalysis system also used a 4D incremental analysis update procedure (4D-IAU, Lei and Whitaker 2016) to reduce gravity-wave noise, and it used the linearized forward operator to compute observation prior ensemble perturbations (Shlyueva and Whitaker 2018).

Analyses were produced four times daily at 00, 06, 12, at 18 UTC, though the reforecasts were initialized at the end of the incremental analysis update window 3 h after the synoptic time. Figure 2 is a schematic illustrating the analysis workflow. This setup differs from the NOAA operational configuration in that (1) the 80-member ensemble was C384 (~25 km) resolution in operations vs. C128 for the reanalysis, and (2) the background "control" forecast used in the operational hybrid 4D-En-Var step came from a deterministic C768 (~13 km) resolution forecast vs. C384 in the reanalysis. After generating the reanalyses, we discovered a bug in the way snow was updated (see section 3d). To partially correct for this, we subsequently 'replayed' (Orbe et al. 2017, section 2.1) the C384 deterministic solution to the C128 ensemble-mean analysis (after recentering the around the hybrid 4D-En-Var analysis) to downscale the C128 analysis to C384 resolution, while updating the land surface states using the correct snow analysis at each analysis time. The replay procedure utilized the same 4D-IAU process used in the data assimilation cycle. The only difference is that instead of re-computing the analysis increments by re-running the hybrid 4D-En-Var data assimilation, it used the previously generated C128 analysis to compute the increments used to constrain the higher resolution model trajectory. This C384 'replay' analysis was used as the ensemble-mean state used to initialize the C384 reforecast ensemble (after superimposing the upscaled C128 ensemble perturbations).

For the 4D-IAU procedure, analyses were produced every 3 h within the 6-h assimilation window (at the beginning, middle, and end of the window). The model was restarted from

the beginning of the assimilation window and forced by analysis increments interpolated in time to the model time step from the 3-hourly fields.

Observation quality control and bias correction was performed by the operational variational Global Statistical Interpolation (GSI) code and was configured as in NCEP operations as of 2015.

c. Stochastic physics in the cycled ensemble for data assimilation.

Model uncertainty in the background ensemble and in the 4D-IAU corrector segment was parameterized by a suite of schemes that consisted of Stochastically Perturbed Physics Tendencies (SPPT, Palmer et al. 2009), Stochastic Kinetic Energy Backscatter (SKEB, Shutts 2005, Berner et al. 2008) and stochastic specific humidity perturbations in the boundary layer (SHUM). The details of the GEFS implementation of SPPT and SKEB are outlined in Zhou et al. (2021).

There were several differences in the model uncertainty parameterizations used in the cycled data assimilation compared to what were used in the GEFSv12. The first difference is that a single spatial and time scale for the random pattern was used in SPPT for the cycled ensemble data assimilation. The background ensemble used an e-folding horizontal length scale of 500 km, an e-folding time scale of 6 h, and an amplitude of 0.8, whereas the GEFSv12 predictions used multiple length and time scales similar to the ECMWF seasonal forecast system (Molteni et al. 2011, Zhou et al. 2021).

In several other respects, the application of the ECMWF SPPT methodology was changed. In early tests to integrate the SPPT to work with the FV3-based GFS, frequent model crashes occurred. These were traced to the interaction of the planetary boundary layer (PBL) scheme and the mountain-blocking scheme. Both of these schemes produced

277 very large momentum tendencies in regions of mountain blocking. When perturbed, they
278 often produced unphysical oscillations that cause numerical stability issues. The solution
279 chosen was to apply SPPT tendencies only above a dividing streamline, which was defined
280 as the level in the atmosphere at which the flow below was blocked due to orography; the
281 SPPT scheme was not active in the portion of the atmosphere that was considered blocked.
282 Another issue with the implementation of SPPT was a wet precipitation bias globally. The
283 perturbed moisture removed from by the microphysics was not reflected in the precipitation
284 at the surface. The solution was to perturb the surface precipitation with the same random
285 number to ensure physical consistency with the perturbed tendencies in the atmosphere.

286 The stochastic kinetic-energy backscatter scheme (SKEB) was the same as GEFSv12,
287 but here we used a length / time scale amplitude of $500 \text{ km } 6 \text{ h}^{-1}$.

288 Additionally, another stochastic scheme was active for the cycled background ensemble
289 that was not used in GEFSv12. Specific humidity tendency perturbations were applied to
290 the lower layers of the model each physics time step. The perturbations rapidly decreased
291 in amplitude in the above the surface, with an e-folding scale of 0.2 in the vertical
292 coordinates (sigma). Thus if the surface pressure was 1000 hPa, the SHUM tendency
293 perturbation at 800 hPa would be ~36.7% of the surface. The humidity perturbations were
294 intended to account for unrepresented variability in small-scale convective triggering
295 features such as gust fronts, cold pools, and sub-grid humidity variability. This stochastic
296 scheme was inspired by Tompkins and Berner (2008), which used random samples of
297 specific humidity distributions provided by the cloud scheme in the convection scheme. As
298 opposed to the Tompkins and Berner methodology which had ensembles of humidity
299 profiles available, in the member-by-member processing of each background ensemble
300 forecast, only the single humidity column profile was available at a given grid point with this

model's parameterization schemes. Hence, our approximation to their algorithm was to perturb the specific humidity with a random value, which is on the order of 0.1% per time-step. Like SPPT and SKEB, the perturbations were correlated in space and time with a timescale of 6 hours and spatial scale of 500 km.

d. Land, ocean, and sea ice initialization.

We produced an atmosphere-only reanalysis to reduce computational expense. Previously generated optimal interpolation (OI; Reynolds et al. 2002) version 2 SST analyses at 1/4 degree provided the ocean state. This differs from the GEFSv12 real-time SST initialization procedure, known as “NSST” (near sea-surface temperature; Minnett et al. 2019). As there can be strong diurnal variations in the ocean skin temperature, especially under sunny conditions and weak winds, the NSST included some simplified dynamics of vertical mixing in the top ocean layers and its variation with atmospheric forcings. The background ocean SST state predicted by NSST is updated to available in-situ and satellite observations. Use of the NSST algorithm was not included in the GEFSv12 reanalysis because its use resulted in excessive SST bias in climatologically cloudy regions during the early years of the reanalysis (not shown), when there were fewer in-situ SST observations and infrequent IR views of the ocean.

There were some systematic differences between SSTs from the OI and from NSST. Figure 4 shows the mean skin temperature at 00 UTC from the pre-production parallel SST analyses minus the skin temperature in the reanalysis during a period from late 2017 to late 2019. Over the ocean, the skin-temperature reanalysis was the OI SST analysis. Generally, mean differences were small over the oceans, though differences in excess of 1C were found in higher-latitude oceans. The lack of major differences over the tropics suggests that surface-based convection should trigger similarly when initialized from

reanalyses as in the real-time production system. This is probably of greater consequence to representation of the general circulation than the mid-latitude differences.

Initialization of snow and ice were designed to be as described in Saha et al. (2010). One known bug in this reanalysis was that the cycled DA erroneously inserted climatological snow amounts for three of the update cycles each day (00, 06, and 12 UTC), while the actual snow analysis was inserted only at 18 UTC. The previously described replay procedure used to re-generate the control C384 analysis using the correct snow analysis at 18 UTC and model-generated snow fields at 00,06 and 12 UTC.

There was no direct data assimilation of top-level soil moisture observations or two-meter temperature/humidity in this reanalysis as there was in other operational prediction facilities (e.g., ECMWF, 2019), though related techniques are in development within NOAA. Further, unlike the real-time GEFSv12, was there no insertion of a standalone soil-state analysis from the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004, Meng et al. 2012), as multi-decadal GLDAS reanalyses were not available at the time of GEFSv12 reanalysis production. Instead, the background forecasts provided precipitation, temperature, and radiative forcings to the underlying cycled land model, but there was no explicit soil-state update based on atmospheric values as there. Deep-soil moisture (layers 2-4) were relaxed to externally specified climatology with a time scale of 60 days, with the climatology specified from an older version of the GLDAS system.

There were some larger skin-temperature differences over land between the reanalysis and the pre-production parallel. In regions with more varied terrain, some of these may represent the differences in the computational grids, which was twice as high in the pre-production parallel (C768) than in the reanalysis (C384), with more terrain detail in the former. Still, areas such as the Himalayan plateau had mean temperatures warmer by

several degrees C in the pre-production parallel analyses. Panels (b) and (c) of Fig. 4 show that the differences around China were persistent from cool to warm season and thus were unlikely due only to the previously mentioned bug related to the snow initialization once per day. Perhaps this region was particularly bias prone, and the insertion of GLDAS states in the pre-production parallel raised these surface temperatures relative to the reanalysis without GLDAS.

4. Statistical characteristics of the reanalysis and of forecasts from them.

a. Conservation properties and the Quasi-Biennial Oscillation.

Figure 5(a) shows the time series of the analyzed global-mean dry surface pressure, which should be constant. There is no explicit constraint during the data assimilation to enforce constancy, and hence the dry surface pressure reflects random and systematic changes to the thermodynamic structure from the assimilation of observations. Analyzed pressure exhibits two non-meteorological jumps, the first around 13-15 July 2006 and the second 11-13 Oct 2009. The first of the pressure jumps coincides with the advent of the assimilation of global positioning system radio occultations (GPSRO; Kursinsky et al. 1997, Anthes et al. 2008). When these unbiased GPSRO observations began to be assimilated in large numbers, they provided an effective anchoring of the microwave radiance data (Cucurull et al. 2014), and these radiance bias corrections changed notably over a short period. The result was that the analyzed thermodynamic structure of the atmosphere changed, and with it the integrated surface pressure. An in-depth examination of the second pressure jump did not reveal any such conclusive cause.

The water cycle in the GEFSv12 reanalysis was not perfectly closed. Fig. 5(b)-(c) shows time series of globally averaged precipitation and evaporation and their difference. This

372 makes the water-cycle balance comparable to the MERRA reanalysis but worse than in
373 MERRA-2. The MERRA-2 reanalysis used extra algorithmic adjustments to ensure near
374 closure of the water cycle (Bosilovich et al. 2017, Fig. 1).

375 Basic characteristics of the analysis of the quasi-biennial oscillation (QBO; Coy et al.
376 2016, Pascoe et al. 2005) are shown in Fig. 6. The QBO diagnostics are based on the
377 monthly and zonally averaged reanalysis u -wind component, -10°S to 10°N . The annual
378 cycle was first removed by subtraction of the 20-year average of each month. Fig. 6(a)
379 provides the winds as a function of pressure (ordinate) and date (abscissa). Filtering was
380 then performed with Fourier analysis, retaining only the first to the 19th harmonics to retain
381 variability longer than a year (periods of 12.6 - 240 months). The filtered data are presented
382 in Fig. 6(b). The time vs. pressure series closely resembles those presented from other
383 reanalyses (e.g., Coy et al. 2016, Fig. 5). The coarser stratospheric resolution in this
384 reanalysis does not permit examination of characteristics above 10 hPa.

385 *b. Fit of the background forecasts to observations.*

386 One way of monitoring the quality of the reanalysis is to display time series of the
387 statistics of differences between the short-term background forecasts (interpolated or
388 converted to the observation type) and the observations. As more observations are added
389 and as the forecast model and assimilation algorithms are improved from one reanalysis
390 system to the next, one would expect the control background forecast to more closely fit the
391 observations in the newer reanalysis system. As an example, the fit of background
392 forecasts to “conventional” observations (primarily rawinsondes, aircraft, and surface and
393 marine in-situ observations) are shown in Fig. 7 for the 800 hPa to 900 hPa layer for both
394 this reanalysis and the previous generation reanalysis, the CFSR. Individual dots denote
395 weekly averaged values, and fitted curves with annual cycles are also plotted. Root-mean-

square errors in this reanalysis were generally decreased relative to CFSR throughout the reanalysis period, while biases were somewhat larger than CFSR in the first decade and smaller in the second decade. Similarly, fits of the background forecasts to the NOAA-15 satellite's Advanced Microwave Sounding Unit A (AMSU-A) channel 8 are shown in Fig. 8. This channel has a peak of its weighting function around 150 hPa. The fit to these observations was notably improved relative to CFSR. In most cases, other levels and other microwave channels provided qualitatively similar improvement of fits. A few satellite channels had, at first glance, had much poorer background fits to the observations in the new reanalysis (not shown). However, these channels assimilated cloudy microwave radiances that the CFSR reanalysis did not use. The cloudy radiances typically have much larger differences between background and observed. In this way, the two reanalyses use different sets of observations, and direct comparisons cannot be easily performed.

c. Scout-run prediction error characteristics.

Another way to evaluate the quality of the reanalysis was to conduct deterministic predictions initialized from the reanalysis and initialized from the reference standard, the previous generation CFSR. Accordingly, for each day during the reanalysis period, five-day deterministic GFS predictions were generated from both reanalyses at C384 resolution. The two temperature predictions were then compared against an independent reference standard, in this case the ERA5 reanalysis (Hersbach et al. 2019). Figure 9 synthesizes the results as a function of the forecast lead time (abscissa) and pressure level (ordinate). The top rows show the time-averaged root-mean-square error (RMSE) of the GEFSv12 predictions minus the RMSE from the CFSR predictions; red colors indicated where the CFSR had lower error relative to the ERA5 reference standard, blue indicated where the GEFSv12 had lower errors. GEFSv12 temperature reanalyses were generally lower in

error, with the notable exception of tropical temperatures just below the tropopause. The GEFSv12 reanalysis used the new FV3 dynamical core as well as a new microphysics parameterization (Zhou et al. 2019). The reanalysis also fully applied humidity observation increments in the stratosphere. In similar testing of the pre-production cycle of GFS data assimilations and forecasts (the “retro” runs), similar humidity biases were noted, and the eventual operational GFS data assimilation configuration was changed to taper the data assimilation humidity increments in the stratosphere to zero. We assume that the upper-tropospheric and lower stratospheric temperature biases were related to this incomplete tuning of the reanalysis system prior to production.

The second row of Fig. 9 shows time-averaged prediction biases of the GEFSv12 reanalysis relative to the ERA5 analyses. The third row of Fig. 9 provides similar plots but for predictions initialized from the CFSR. Temperature biases were more pronounced in the forecasts initialized from the GEFSv12 reanalysis. Cold biases near 250 hPa increased with lead time and were especially prominent in the tropics. There were also more pronounced warm stratospheric biases in predictions from the GEFSv12 reanalysis. While lower in magnitude, the increasing biases with lead time in CFSR-initialized forecasts indicates that the bias originated in the version of the FV3-basex GFS forecast model used here, not in some aspect of the data assimilation.

Did the apparently increased upper-air temperature bias in the GEFSv12 result in forecasts with degraded accuracy? Figure 10 shows anomaly correlation (Wilks 2011, section 8.6.4) “dieoff curves” for GFS deterministic predictions of temperature initialized from the GEFSv12 system (red) vs. from the CFSR (dashed blue). Five lines are plotted, one each representing the average over one of the five streams. While there was some variability in performance from one stream to the next (generally, the lower anomaly

correlations correspond to earlier streams with sparser observational data), consistently across levels and regions, the predictions initialized from the GEFSv12 system were higher in anomaly correlation, especially in the tropics despite the temperature bias. This reinforces the inferences from fits to observations (Figs. 7, 8) that the GEFSv12 reanalysis had reduced errors relative to CFSR.

d. Comparisons with pre-production parallel forecasts.

In addition to the multi-decadal GEFSv12 reanalyses and the reforecasts generated from them, NOAA partners at the Environmental Modeling Center also performed approximately two years of retrospective cycled data assimilation and GEFSv12 forecasts with approximately the model and assimilation system version and that were made operational, at the target operational resolution. These “retro” forecast data were primarily used as a comparison against the then-operational system to ensure that the new forecast system produced forecasts of equal or greater quality. The retro forecasts form a useful baseline for comparison of the reforecasts, which, it was hoped, would be similar in error characteristics. The cycled “retro” assimilations and forecasts did not use a strictly unchanging model and data assimilation configuration; as biases were discovered, parameter settings were changed. In this way, it proved difficult to isolate the specific parameter settings that may have been responsible for systematic differences between retro and reforecasts beyond the obvious differences in analysis resolution.

Figure 11 illustrates differences between global 5-member ensemble-mean reforecasts and the mean of five members of the retro ensemble forecasts. The top row compares the mean prediction characteristics in the absence of any verifying reanalysis, the mean retro minus mean reforecast. As shown in Figs. 9 (a) - (c), Fig. 11(a) indicates that the reanalysis-initialized predictions had mean temperatures that were lower than those from

the retro runs near 250 hPa, and reforecast-initialized predictions of the stratosphere were warmer. Reforecast- initialized predictions of humidity above 200 hPa were lower than retro-run predictions. Temperature biases, through geostrophic adjustment, presumably affected the wind structure above this level. Most likely, these systematic differences were primarily attributable to differences in the cloud microphysical parameterization used in the cycled reanalysis and the application of a tapering of humidity increments in the stratosphere in the retro's cycled data assimilation. These tapering changes had not been made to the reanalysis system at the time of the production. Retro-run cycled assimilations and predictions were performed later, after subsequent tuning of the microphysics reduced biases and application of tapering.

A curious feature is shown in Fig. 11(i), where u -component differences in the 200-400 hPa layer grow through day +7 but then get smaller. This can be explained by the differing lead times of forecast-error saturation. Generally, reforecast-initialized winds at these levels had forecast errors corresponding to a 6-h loss in prediction time relative to the retro, i.e., reforecast-initialized forecasts at +7 days lead were as accurate as retro-run initialized forecasts at +7.5 days. The global saturation time scale was largely controlled by the (shorter) value in the tropics (not shown), given the large fraction of the earth's surface area that is between 30° S and 30° N. Thus, global reforecast prediction errors were more quickly reaching a saturation error value than retro predictions. But the retro predictions eventually began to saturate shortly thereafter as well. Despite different reforecast vs. retro errors and biases in upper-air fields, the differences were smaller near the earth's surface, which was of greater concern for most forecast applications such as postprocessing of sensible weather variables.

Another way of evaluating the relative quality of forecasts initialized from the reanalysis and from the retro runs is with anomaly correlation dieoff curves (Fig. 12), as in Fig. 10. There were differences in the anomaly correlations of the reforecast-initialized forecasts, which were slightly lower than for the retro runs. Consistent with Fig. 11, these were larger at higher levels and modest at lower levels.

5. Known discrepancies between the reanalysis and the real-time analysis.

Aside from the obvious resolution differences, Table 2 synthesizes what we believe to be the major discrepancies between the configuration of the reanalysis and the real-time forecast configuration.

6. Description of reanalysis data storage.

Six-hourly control reanalyses for the 2000-2019 period are publicly available on the “emcrzdm” server. 590 variables at 0.25 degrees were created for the synoptic times 00, 06, 12 and 18 UTC in grib2 format. Because this server also provides operational forecast data, for some times of day, download speeds may be slow. The data can be found here: ftp://ftp.emc.ncep.noaa.gov/GEFSv12/reanalysis/FV3_reanalysis/

7. Conclusions.

This article has described the algorithms behind and the characteristics of a multi-decadal global ensemble atmosphere-only reanalysis covering 2000-2019. It was designed for initializing the atmospheric component of reforecasts for the new US National Weather Service Global Ensemble Forecast System, version 12 (GEFSv12), described in a companion article (Guan et al. 2021). It has approximate consistency with the operational data assimilation procedure used to initialize the real-time GEFSv12 forecasts. Partly

513 because of computational expense that necessitated lower reanalysis resolution and partly
514 because the reanalysis had to be created before the operational configuration was settled,
515 there were inconsistencies. For one, reanalysis resolution was lower. Also, upper-air
516 thermodynamic variables were different between reanalyses and real-time analyses due to
517 different parameterization configurations, and that affected the upper-air wind analyses as
518 well. There were some near-surface differences, in one case due to a bug in the updating
519 to snow analyses, as well as to different real-time vs. reanalysis procedures for land and
520 SST initialization.

521 We encountered significant challenges in the production of this reanalysis that are
522 motivating some potential changes the next time a NOAA modern-era reanalysis is created.
523 The most significant problem was that reanalysis computations had to be spread over a long
524 period of time due to NOAA's supercomputers being saturated. Consequently, the
525 reanalysis configuration was decided roughly one year before final decisions were made on
526 the real-time analysis configuration used for GEFSv12 initialization, and computations were
527 spread over a long period. Perhaps projects like reanalyses and the accompanying
528 reforecasts are more appropriate for *cloud computing*, where their computations can be
529 delayed until just before a model implementation and then performed over a short period of
530 time. With the short production period, a consistent system configuration can be used for
531 both reanalysis and the real-time system. The potential downside of this approach is the
532 expense of cloud computing. However, when factoring in all the comparative costs of
533 NOAA-owned vs. cloud high-performance computing, the costs may be more comparable,
534 especially if cloud computations are performed with less expensive "spot" instances.

535 Data preparation and quality control is always a challenge with reanalyses. While we
536 leveraged archived NOAA observational data used in a previous reanalysis, the international

537 sharing of reanalysis observational data sets would have many desirable consequences.
538 With more people involved in data preparation and quality control, each organization would
539 have greater confidence in the fidelity of these data. Also, were standardized observation
540 data sets used by multiple organizations, then reanalysis comparisons will become more
541 straightforward, and differences in quality can be attributed to model and assimilation
542 system design rather than to observation data-set differences.

543 Particular challenges in consistency of reanalysis vs. real-time analysis also occurred
544 near the earth's surface. Consistent global land-data assimilation system (GLDAS)
545 reanalysis states were not available at the time of reanalysis production, and the real-time
546 sea-surface temperature initialization procedure did not perform acceptably in the more
547 data-sparse environment early in the reanalysis period. Hence, the development of
548 consistent, coupled ocean-atmosphere data assimilation procedures shared between the
549 reanalysis and the real-time system are a priority. Similarly, improvement of land-data
550 assimilation procedures so that they can be implemented consistently between reanalysis
551 and real-time operations are also desirable. NOAA is currently working on procedures that
552 improve on GLDAS procedures by using 2-m temperature and humidity observations and
553 other data to make increments to the soil state (Draper 2021). When implemented across
554 applications, this should both improve the consistency and make future reanalyses more
555 widely useful, such as for climate monitoring applications in addition to reforecast
556 initialization.

557 **Acknowledgments:**

558 Jonathan Brannock of NOAA's Big Data Project provided assistance in moving data to
559 Amazon Web Services. Zac Lawrence of PSL and CIRES performed the analysis of the
560 Quasi-Biennial Oscillation in this paper. This project was supported by NOAA Weather

561 Program Office grant T8R1CRS-PCR, NOAA Climate Program Office grants S8R1CWS-
562 PRE and R8R1CWS-P01, and NWS OSTI grants P8MWQNG-PTR, P8R1MP1-PPS, and
563 M8M6H81-P10.

References

- Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 Mission: Early Results. *Bull. Amer. Meteor. Soc.*, **89**, 313–334, <https://doi.org/10.1175/BAMS-89-3-313>.
- Berner J., G. Shutts, M. Leutbecher, and T.N. Palmer, 2008: [A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system](#). *J. Atmos. Sci.*, **66**, 603–626.
- Bosilovich, M.G., F.R. Robertson, L. Takacs, A. Molod, and D. Mocko, 2017: [Atmospheric water balance and variability in the MERRA-2 reanalysis](#). *J. Climate*, **30**, 1177–1196, <https://doi.org/10.1175/JCLI-D-16-0338.1>
- Browning, K.A., and R. Wexler, 1968: The determination of kinematic properties of a wind field using Doppler radar. *J. Appl. Meteor.*, **8**, 105–113.
- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley, S.J., 2011: The Twentieth century reanalysis project. *Quart. J. Royal Meteor. Soc.*, **137**, 1–28. doi:[10.1002/qj.776](https://doi.org/10.1002/qj.776)
- Courtier, P., Thépaut, J.-N. and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Royal Meteor. Soc.*, **120**, 1367–1387. doi:[10.1002/qj.49712051912](https://doi.org/10.1002/qj.49712051912)
- Coy, L., K. Wargan, A.M. Molod, W.R. McCarty, and S. Pawson, 2016: [Structure and dynamics of the Quasi-Biennial Oscillation in MERRA-2](#). *J. Climate*, **29**, 5339–5354, <https://doi.org/10.1175/JCLI-D-15-0809.1>
- Cucurull, L., R. A. Anthes, and L. Tsao, 2014: Radio occultation observations as anchor observations in numerical weather prediction models and associated reduction of bias

corrections in microwave and infrared satellite observations. *J. Atmos. Oceanic Technol.*, **31**, 20–32, <https://doi.org/10.1175/JTECH-D-13-00059.1>.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Royal Meteorol. Soc.*, **137**, 553-597. doi:[10.1002/qj.82](https://doi.org/10.1002/qj.82)

Draper, C., 2021: Accounting for land model uncertainty in numerical weather prediction ensemble systems: toward ensemble-based coupled land/atmosphere data assimilation, *J. Hydrometeor.*, submitted.

ECMWF, 2019: IFS Documentation—Cy46r1 operational implementation 6 June 2019. Part II: Data assimilation. ECMWF, 8 Dec 2019. Available from <https://www.ecmwf.int/en/elibrary/19306-part-ii-data-assimilation>.

Gelaro, R., R.H. Langland, S. Pellerin, and R. Todling, 2010: [The THORPEX Observation Impact Intercomparison Experiment](https://doi.org/10.1175/2010MWR3393.1). *Mon. Wea. Rev.*, **138**, 4009–4025, <https://doi.org/10.1175/2010MWR3393.1>

Gelaro, R., W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G. Kim, R. Koster, R. Lucchesi, D. Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, and B. Zhao, 2017: [The Modern-Era Retrospective Analysis](https://doi.org/10.1175/2017MWR3393.1)

[for Research and Applications, Version 2 \(MERRA-2\)](#). *J. Climate*, **30**, 5419–5454,
<https://doi.org/10.1175/JCLI-D-16-0758.1>

Guan, H., and others, 2021: The GEFSv12 reforecast dataset for supporting sub-seasonal and hydrometeorological applications. In preparation.

Hagedorn, R, T. M. Hamill, and J. S. Whitaker, 2008: [Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: 2-meter temperature](#). *Mon. Wea. Rev.*, **136**, 2608-2619.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: [Ensemble re-forecasting: improving medium-range forecast skill using retrospective forecasts](#). *Mon. Wea. Rev.*, **132**, 1434-1447.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: [Reforecasts, an important dataset for improving weather predictions](#). *Bull. Amer. Meteor. Soc.*, **87**, 33-46.

Hamill, T. M., and J. S. Whitaker, 2006: [Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application](#). *Mon. Wea. Rev.*, **134**, 3209-3229.

Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: [Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: precipitation](#). *Mon. Wea. Rev.*, **136**, 2620-2632.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y. Zhu, and W. Lapenta, 2013: [NOAA's second-generation global medium-range ensemble reforecast data set](#). *Bull. Amer. Meteor. Soc.*, **94**, 1553-1565.

Hamill, T. M., 2017: [Changes in the systematic errors of global reforecasts due to an evolving data assimilation system](#). *Mon. Wea. Rev.*, **145**, 2479-2485.

Harada, Y., H. Kamahori, C. Kobayashi, H. Endo, S. Kobayashi, Y. Ota, H. Onoda, K.

Onogi, K. Miyaoka, and K. Takahashi, 2016: The JRA-55 Reanalysis: Representation of

atmospheric circulation and climate variability, *J. Meteor. Soc. Japan*, **94**, 269-302,
doi:10.2151/jmsj.2016-015.

Helfrich, S. R. , D. McNamara , B. H. Ramsay, T. Baldwin, and T. Kasheta, 2007:
Enhancements to, and forthcoming developments in the Interactive Multisensor Snow
and Ice Mapping System (IMS). *Hydrol. Processes*, **21**, 1576–1586,
doi:10.1002/hyp.6720.

Hersbach, H., B. Bell, P. Berrisford, A. Horányi, J. M. Sabater, J. Nicolas, R. Radu, D.
Schepers, A. Simmons, C. Soci, D. Dee, 2019: Global reanalysis: goodbye ERA-Interim,
hello ERA5. *ECMWF Newsletter* 159, 17-24. Available at
<https://www.ecmwf.int/en/elibrary/19001-newsletter-no-159-spring-2019>

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha,
G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C.
Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996:
[The NCEP/NCAR 40-Year Reanalysis Project](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2). *Bull. Amer. Meteor. Soc.*, **77**, 437–472,
[https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. Yang, J.J. Hnilo, M. Fiorino, and G.L. Potter,
2002: [NCEP–DOE AMIP-II Reanalysis \(R-2\)](https://doi.org/10.1175/BAMS-83-11-1631). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644,
<https://doi.org/10.1175/BAMS-83-11-1631>

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009:
Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea.*
Forecasting, **24**, 1691–1705.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble
data assimilation for the NCEP GFS. Part II: 4D EnVar and hybrid variants. *Mon. Wea.*
Rev., **143**, 452–470, doi:<https://doi.org/10.1175/MWR-D-13-00350.1>.

Kleist, D. T., R. Mahajan, and C. Thomas, 2018: Data assimilation in the Next-Generation Global Prediction System (NGGPS) Era: Initial implementation of FV3-based Global Forecast System (GFS). *JCSDA Quarterly*, **61**, 1-9, doi:10.25923/jw00-r987

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi , 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5-48, doi:10.2151/jmsj.2015-001.

Koster, R. D., and others, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138-1140 DOI: 10.1126/science.1100217

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R., 1997: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, *J. Geophys. Res.*, **102**(D19), 23429–23465, doi:[10.1029/97JD01569](https://doi.org/10.1029/97JD01569).

Laloyaux, P., E. de Boisseson, M. Balmaseda, J.-R. Bidlot, S. Broennimann, R. Buizza, 2018: CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, **10**, 1172– 1195. <https://doi.org/10.1029/2018MS001273>

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi , 2015: The JRA-55 Reanalysis: general specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5-48, doi:10.2151/jmsj.2015-001.

Lee, W., Tang, X., & Jou, B. J., 2014: Distance Velocity–Azimuth Display (DVAD)—new interpretation and analysis of doppler velocity. *Mon. Wea. Rev.*, **142**(2), 573-589. <https://journals.ametsoc.org/view/journals/mwre/142/2/mwr-d-13-00196.1.xml>

- Lei, L., and J. S. Whitaker, 2016: A four-dimensional incremental analysis update for the ensemble Kalman filter. *Mon. Wea. Rev.*, **144**, 2605–2621, <https://doi.org/10.1175/MWR-D-15-0246.1>.
- Lei, L., Whitaker, J. S., & Bishop, C., 2018: Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *Journal of Advances in Modeling Earth Systems*, **10**, 3221– 3232. <https://doi.org/10.1029/2018MS001468>
- Lin, S.-J., 2004: [A “vertically Lagrangian” finite-volume dynamical core for global models.](#) *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2)
- Liu, Q., T. Marchok, H. Pan, M. Bender, and S. Lord, 1999: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. *NOAA Tech. Proc. Bull.* 472, National Weather Service, Office of Meteorology, Silver Spring, MD, 1–7. [Available online at www.nws.noaa.gov/om/tpb/472.htm.]
- Meng, J., Yang R., Wei H., Ek M., Gayno G., Xie P., and Mitchell K., 2012: The land surface analysis in the NCEP climate forecast system reanalysis. *J. Hydrometeor.*, **13**, 1621–1630, doi:10.1175/JHM-D-11-090.1.
- Minnett, P. J., A. Alvera-Azcarate, T.M. Chin, G.K. Corlett, C.L. Gentemann, I. Karagali, X. Li, A. Marsouin, S. Marullo, E. Maturi, R. Santoleri, S. Saux Picart, M. Steele, J. Vazquez-Cuervo, 2019: Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment*, **223**, <https://doi.org/10.1016/j.rse.2019.111366>.
- Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer and F. Vitart, 2011: The new ECMWF seasonal forecast system (System 4). *ECMWF Tech. Memo.* **656**, 1-49.

- Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., & Molod, A. M. (2017). Large-scale atmospheric transport in GEOS replay simulations. *Journal of Advances in Modeling Earth Systems*, **9**, 2545–2560.
<https://doi.org/10.1002/2017MS001053>
- Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A. Weisheimer, 2009: Stochastic Parametrization and Model Uncertainty. *ECMWF Tech. Memo.* **598**, 1-42.
- Pascoe, C. L., Gray, L. J., Crooks, S. A., Juckes, M. N., and Baldwin, M. P., 2005: The quasi-biennial oscillation: analysis using ERA-40 data, *J. Geophys. Res.*, **110**, D08105, doi:[10.1029/2004JD004941](https://doi.org/10.1029/2004JD004941).
- Phillips, V.T.J. and Donner, L.J., 2006: Cloud microphysics, radiation and vertical velocities in two- and three-dimensional simulations of deep convection. *Quart. J. Royal Meteor. Soc.*, **132**, 3011-3033. doi:[10.1256/qj.05.171](https://doi.org/10.1256/qj.05.171)
- Putman, W. M., S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Computational Phys.*, **227**, 55-78 <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Royal Meteor. Soc.*, **126**, 1143-1170.
doi:[10.1002/qj.49712656415](https://doi.org/10.1002/qj.49712656415)
- Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609-1625.
- Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, and J. Woollen,

2011: [MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications](#). *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, 2004: The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, **85**, 381-394, 2004.

Saha, S., S. Moorthi, H. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y. Hou, H. Chuang, H.H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R.W. Reynolds, G. Rutledge, and M. Goldberg, 2010: [The NCEP Climate Forecast System Reanalysis](#). *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>

Scheuerer, M., and T. M. Hamill, 2015: Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted Gamma distributions. *Mon. Wea. Rev.*, **143**, 4578-4596. Also appendix A and appendix B and appendix C.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, 131, 3079–3102. <https://doi.org/10.1256/qj.04.106>

Shlyueva, A., & Whitaker, J. S., 2018: Using the linearized observation operator to calculate observation space ensemble perturbations in ensemble filters. *J. Advances in Modeling Earth Systems*, **10**, 1414– 1420. <https://doi.org/10.1029/2018MS001309>

Slivinski, L.C., Compo, G.P., Whitaker, J.S., and others, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J Royal Meteor. Soc.*, **145**, 2876– 2908. <https://doi.org/10.1002/qj.3598>

Tompkins, A. M., and Berner, J., 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability, *J. Geophys. Res.*, **113**, D18101, doi:10.1029/2007JD009284.

Uppala S.M., and others, 2005. The ERA-40 re-analysis. *Quart. J. Royal Meteorol. Soc.*, **131**, 2961 – 3012.

Vannitsem, S., D. S. Wilks, and J. W. Messner, eds., 2018: *Statistical postprocessing of ensemble forecasts*. Elsevier Press, 347 pp.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for the NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, doi:<https://doi.org/10.1175>

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences (Third Edition)*. Academic Press, 676 pp.

Zhao, Q. and F.H. Carr, 1997: [A Prognostic Cloud Scheme for Operational NWP Models](https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2). *Mon. Wea. Rev.*, **125**, 1931–1953, [https://doi.org/10.1175/1520-0493\(1997\)125<1931:APCSFO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2)

Zhou, L., S. Lin, J. Chen, L.M. Harris, X. Chen, and S.L. Rees, 2019: [Toward Convective-Scale Prediction within the Next Generation Global Prediction System](https://doi.org/10.1175/BAMS-D-17-0246.1). *Bull. Amer. Meteor. Soc.*, **100**, 1225–1243, <https://doi.org/10.1175/BAMS-D-17-0246.1>

Zhou, X., and others 2021: The Introduction of the NCEP Global Ensemble Forecast System Version 12, in preparation.

781 **Table 1:** A synthesis of the major differences between the CFSR and GEFSv12 reanalyses.

Aspect changed	CFSR configuration	GEFS v12 configuration
Period of record	1978-current	2000 - 2020
Atmospheric dynamical core and control forecast grid spacing	Spectral, T382L64 (~ 38 km grid)	FV3 (Lin 2004, Putman and Lin 2007), C384L64 (~ 25 km grid)
Microphysical parameterization	Zhao-Carr (Zhao and Carr 1997)	GFDL (Phillips and Donner 2006, Zhou et al. 2019)
Other parameterizations	Saha et al. (2010)	GFSv15 (2020)
Atmospheric data assimilation methodology	3D-Var (Parrish and Derber 1992, Kleist et al. 2009)	Hybrid En-Var with replay
Ensemble usage in data assimilation	None	80-member EnKF at C128L64 (~ 75 km) to provide background- error covariances.
Ensemble stochastic physics	None (single control member for data assimilation)	Stochastically perturbed physical tendencies (SPPT), stochastic boundary-layer relative humidity (SHUM), and stochastic kinetic-energy backscatter (SKEB) [this paper]
Snow updates	SNODEP (Kiess and Kopp, 1997) before 1997, NESDIS IMS (Helfrich et al. 2007) thereafter. Updated 4x daily.	NESDIS IMS (Helfrich et al. 2007). Updated only at 00 UTC, otherwise climatology for other 3 cycles (a bug).
Land-surface analysis	Separate land-surface analysis with analyzed forcings (Saha et al. 2010)	Land-surface forcings directly from short-term forecasts.

Ocean analysis	SST via OI (Reynolds et al. 2007); rest of ocean state with 3D-Var using MOM4 ocean and weak coupling	SST via OI (Reynolds et al. 2007). No weak coupling in cycled DA, no full ocean analysis.
Tropical cyclone processing	Vortex relocation to observed position (Liu et al. 1999)	Direct assimilation of central pressure, no relocation

782

783

Table 2: Consistency issues in the GEFSv12 reanalysis relative to the operational data assimilation scheme, and their consequences.

Issue	Consequence
OI SST in reanalysis vs. NSST in operations	Reanalysis ocean initial SSTs are systematically warmer in some extratropical areas, with SST reforecasts inheriting this bias. See Fig. 4(a).
Snow analyses only inserted one cycle of four.	Despite replay procedure to ameliorate differences, some land-surface skin temperature bias. See Figs. 4(b) - (c)
Soil temperature and moisture initialization not via GLDAS in reanalysis	Freely cycling land-surface state in reanalysis procedure can result in initial systematic differences of soil moisture and temperature.
Microphysics and parameterization tuning	Biases in upper tropospheric and stratospheric variables in the reanalysis relative to real-time system. See Figs. 9, 11.
US Doppler radar velocity-azimuth display winds turned off	Degraded lower-tropospheric wind analyses prior to their discovery (section 2 of this manuscript)
Velocity azimuth display winds not assimilated in much of reanalysis because of insufficient quality control	Lower-tropospheric wind observations in reanalysis over US somewhat sparser than in real-time analysis.

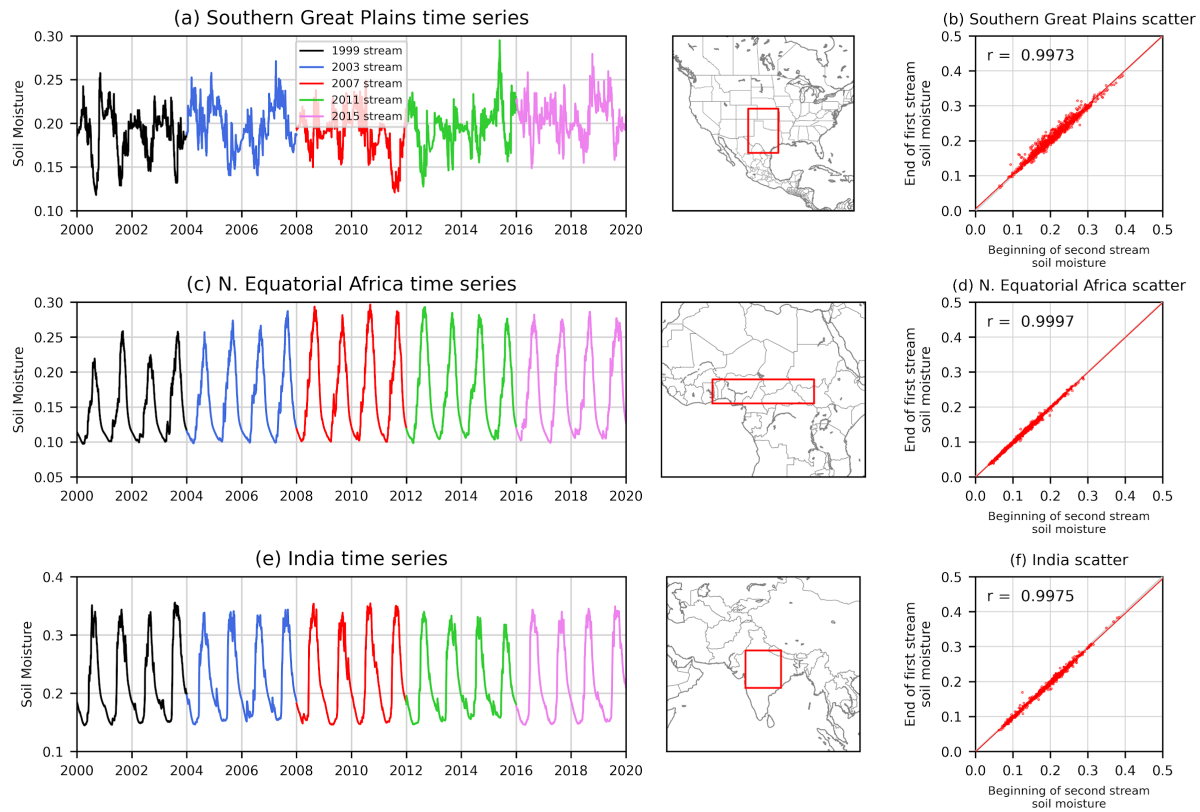


Figure 1: Time series of mean volumetric soil moisture in overlapping streams for three areas with strong land-atmosphere coupling. (a) US Great Plains, (c) Northern equatorial Africa, and (e) India. Also plotted are scatterplots of the soil moistures in each region between the last day of one stream and the first day of the next stream; each dot represents a separate grid point at one of the four stream boundaries.

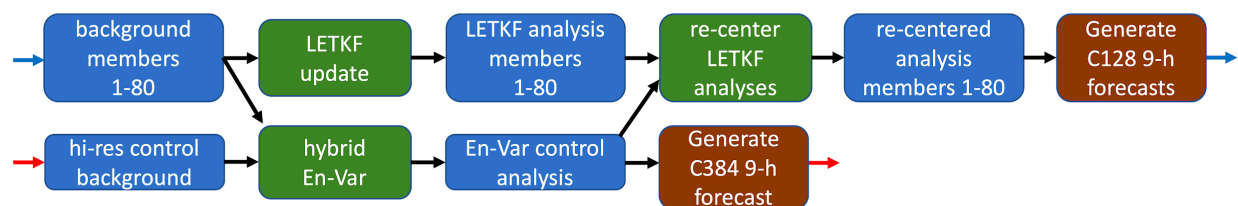
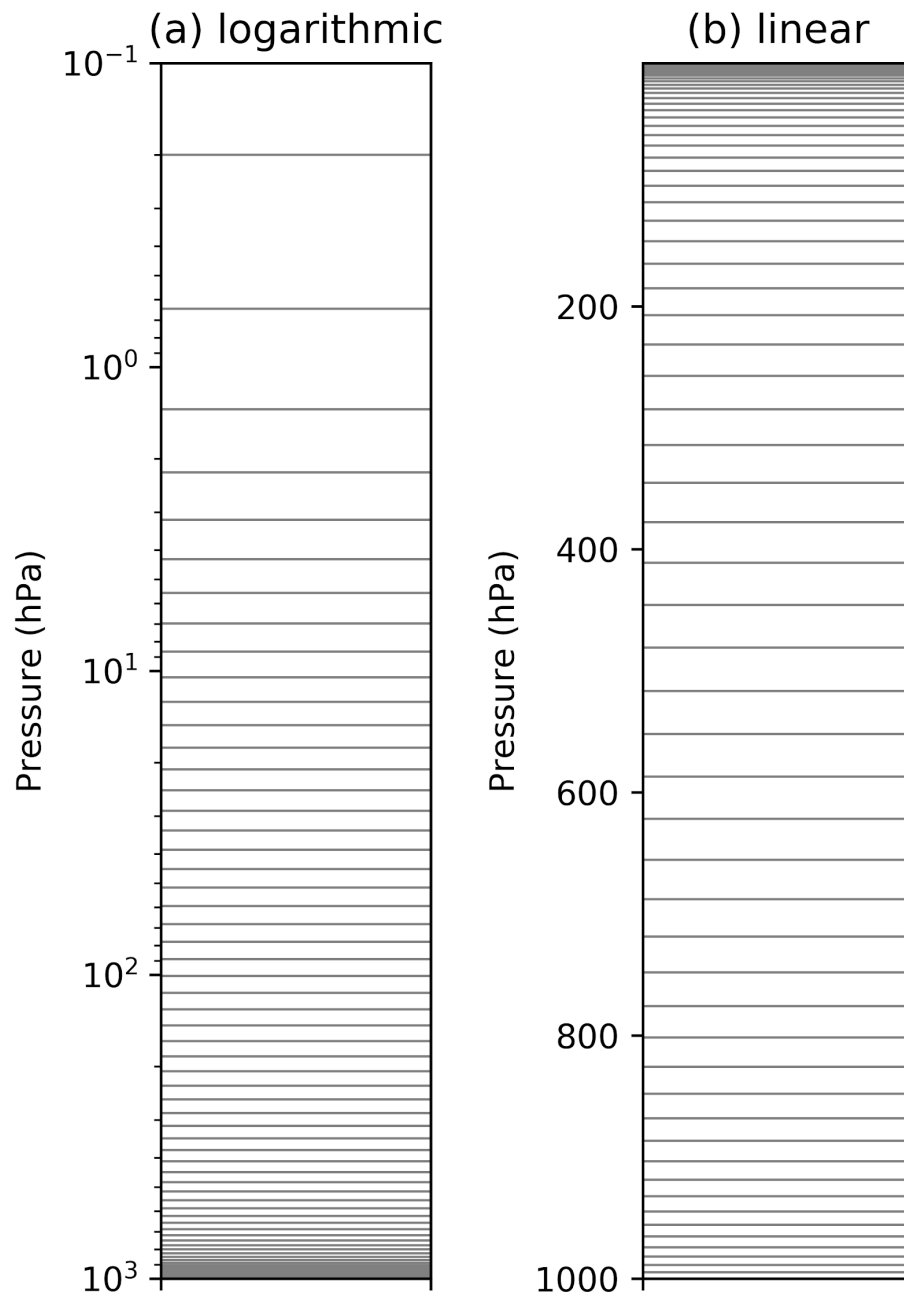


Figure 2: Cycled data assimilation data and process flow used in the GEFSv12 reanalysis.

Cycled lower-resolution LETKF analyses were maintained alongside the higher-resolution control En-Var analysis. Background cycled LETKF forecasts informed the background-error covariance model in the hybrid En-Var. LETKF analyses were re-centered around the En-Var control. The LETKF analyses were computed at ensemble resolution C128, three times lower than operations. The control analysis was computed at C384, two times lower resolution than operations.

Vertical levels



803

804 **Figure 3:** Vertical levels for the cycled data assimilation and forecast model used in this
805 reanalysis for a surface pressure of 1000 hPa, plotted on (a) logarithmic in pressure, and (b)
806 linear in pressure.

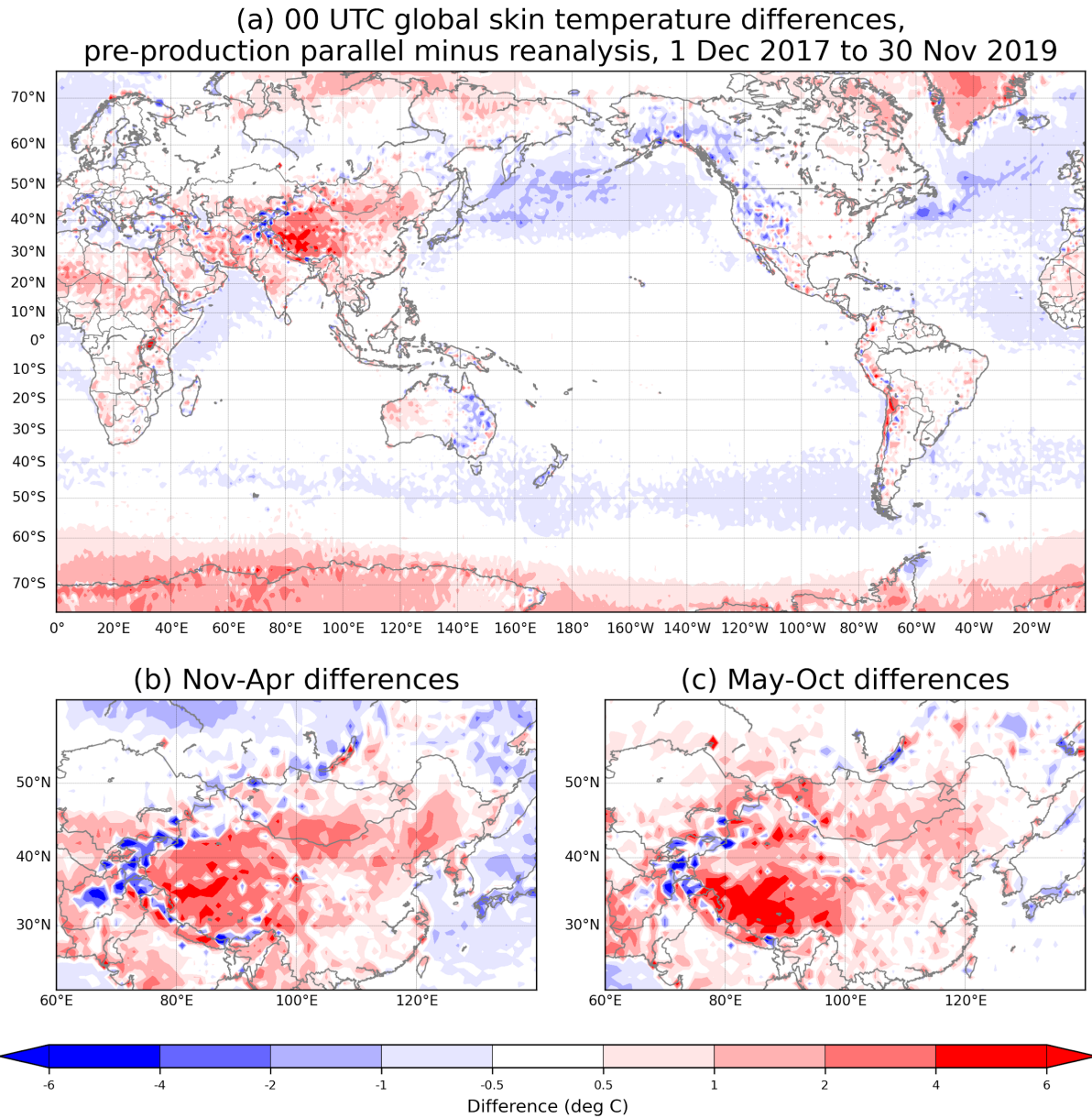


Figure 4: Differences between 00 UTC analyzed skin temperatures, pre-production parallel analysis minus reanalysis, (a) for the period 1 Dec 2017 to 30 November 2019, (b) a subset of the domain for dates within this period and for the months November to April, and (c) for dates within this period from May to October.

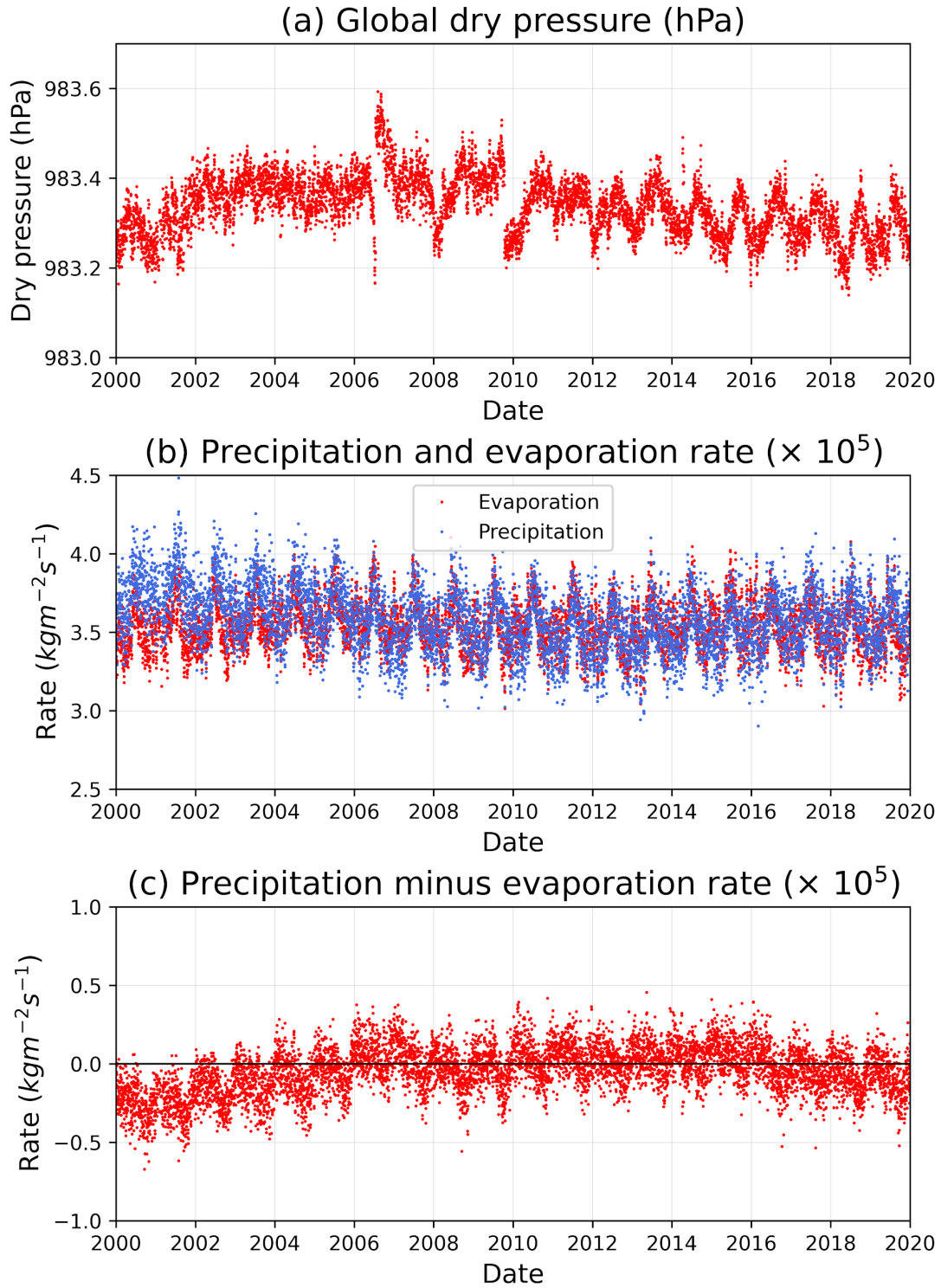
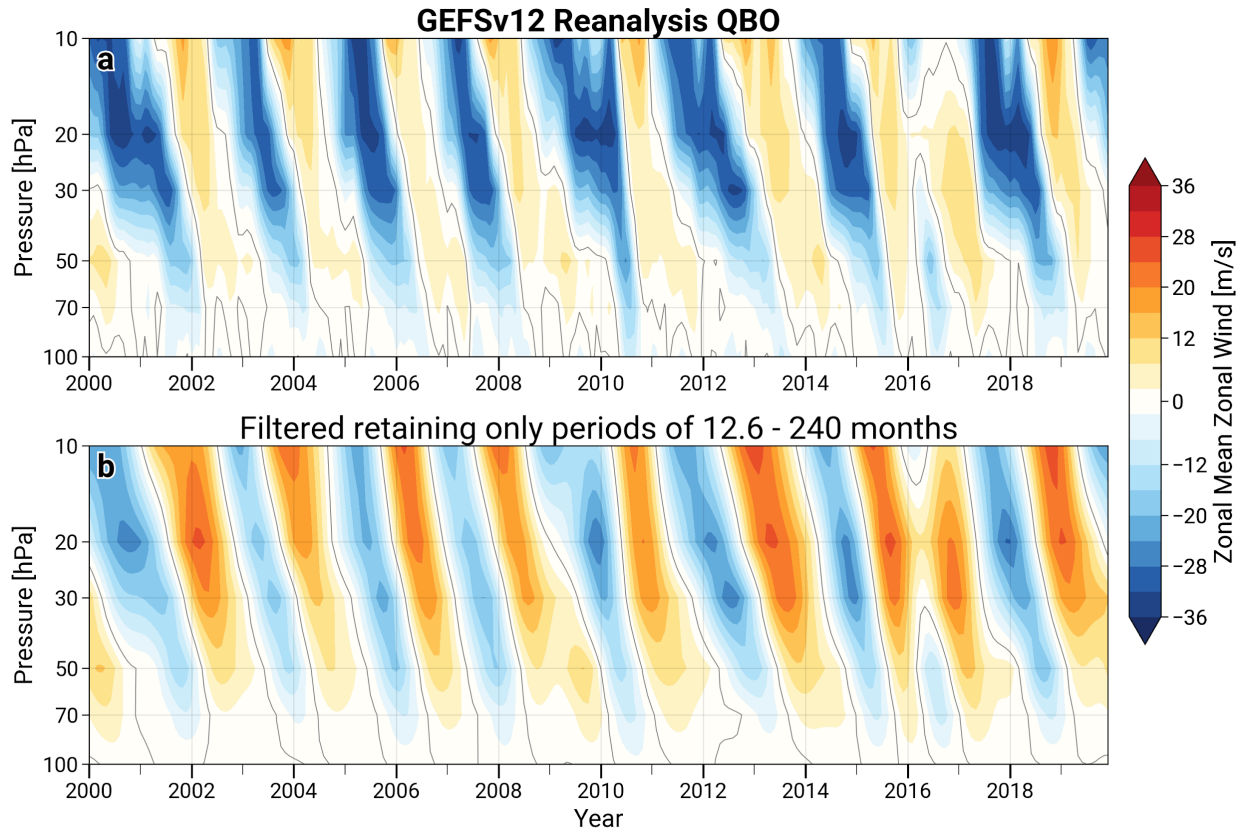


Figure 5: Time series of (a) dry surface pressure, (b) precipitation and evaporation rate, and (c) precipitation minus evaporation rate.



815

816 **Figure 6.** Illustration of the monthly zonal-mean wind in the 10° S to 10° N latitude band,

817 illustrating the QBO in this reanalysis. (a) before filtering, and (b) after.

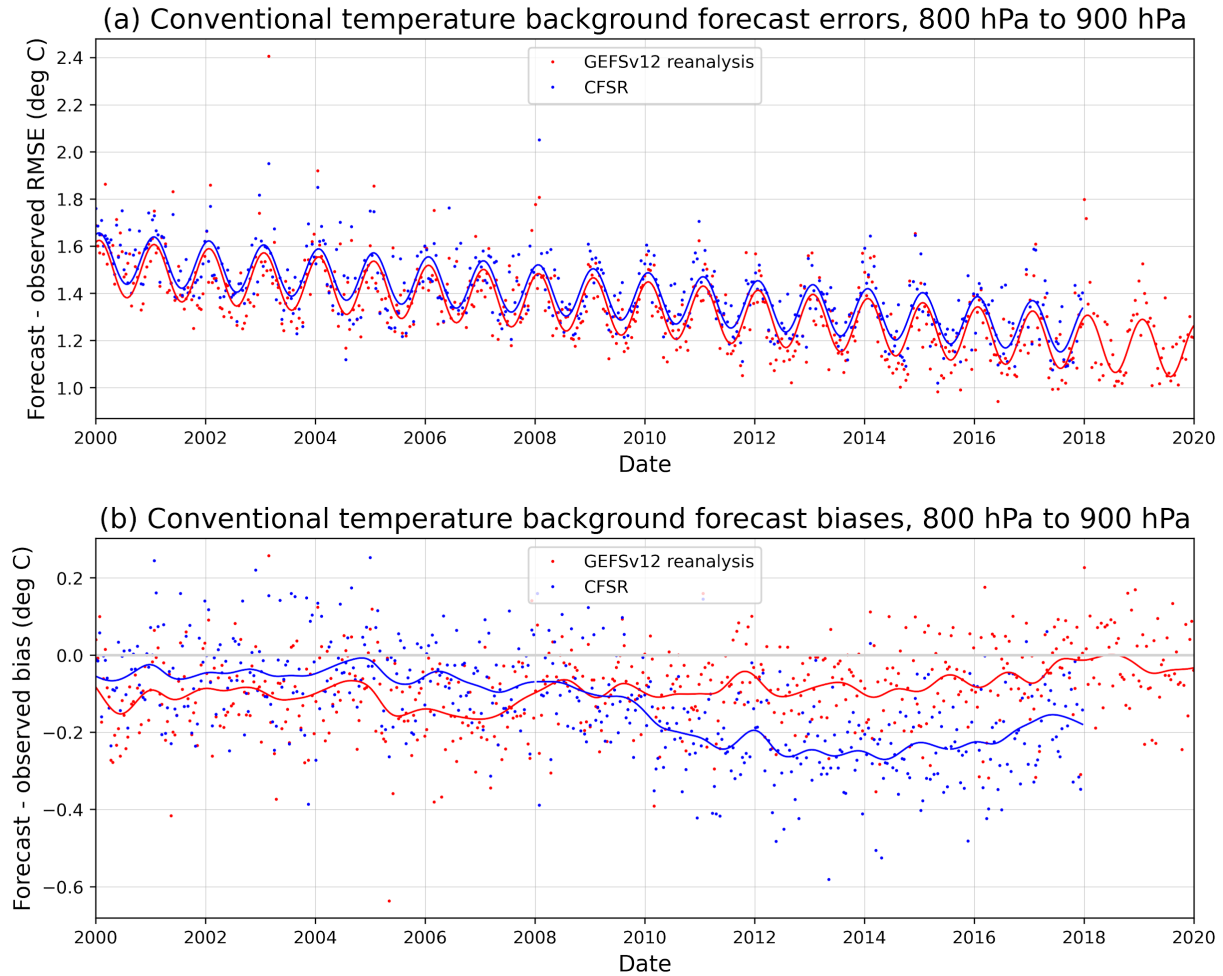


Figure 7: (a) 12.5-day average RMSE fit (dots) of control background forecast to conventional data, red for GEFSv12 analysis and blue for CFSR. (b) averages of control background forecast bias. Colored lines in top panel are estimated with linear regression including an annual cycle, and the lines in the bottom panel are estimated with a Gaussian kernel smoother with a 125-day e-folding time scale.

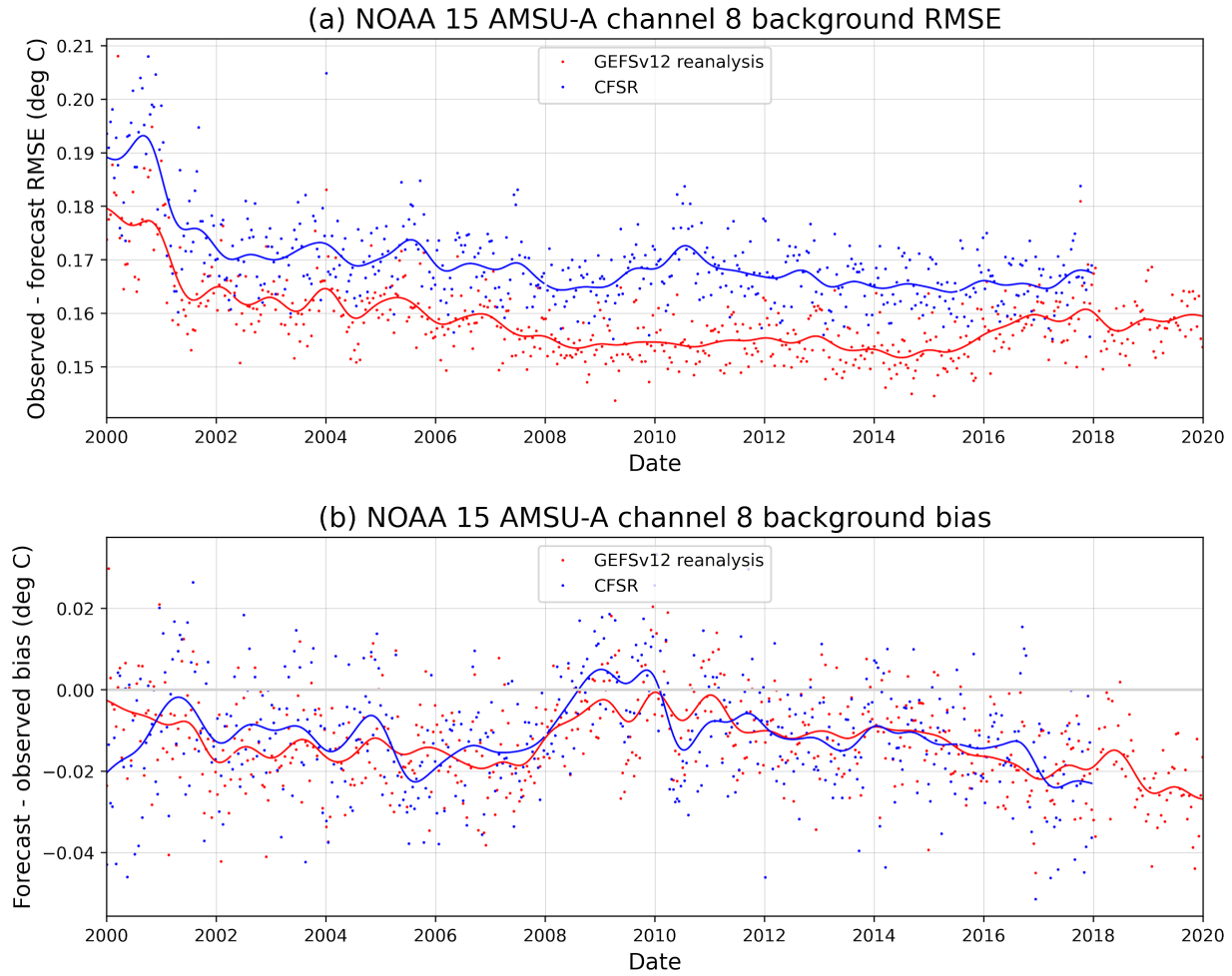


Figure 8: Two-week average RMSE fit (dots) of control background forecast to the NOAA-15 polar-orbiting satellite's AMSU-A channel 8 (peak weighting at ~ 150 hPa), red for GEFSv12 analysis and blue for CFSR. Overplotted are Gaussian kernel smoother RMSEs with an e-folding correlation time scale of 125 days.

2000-2019 temperature forecast RMSE difference and biases

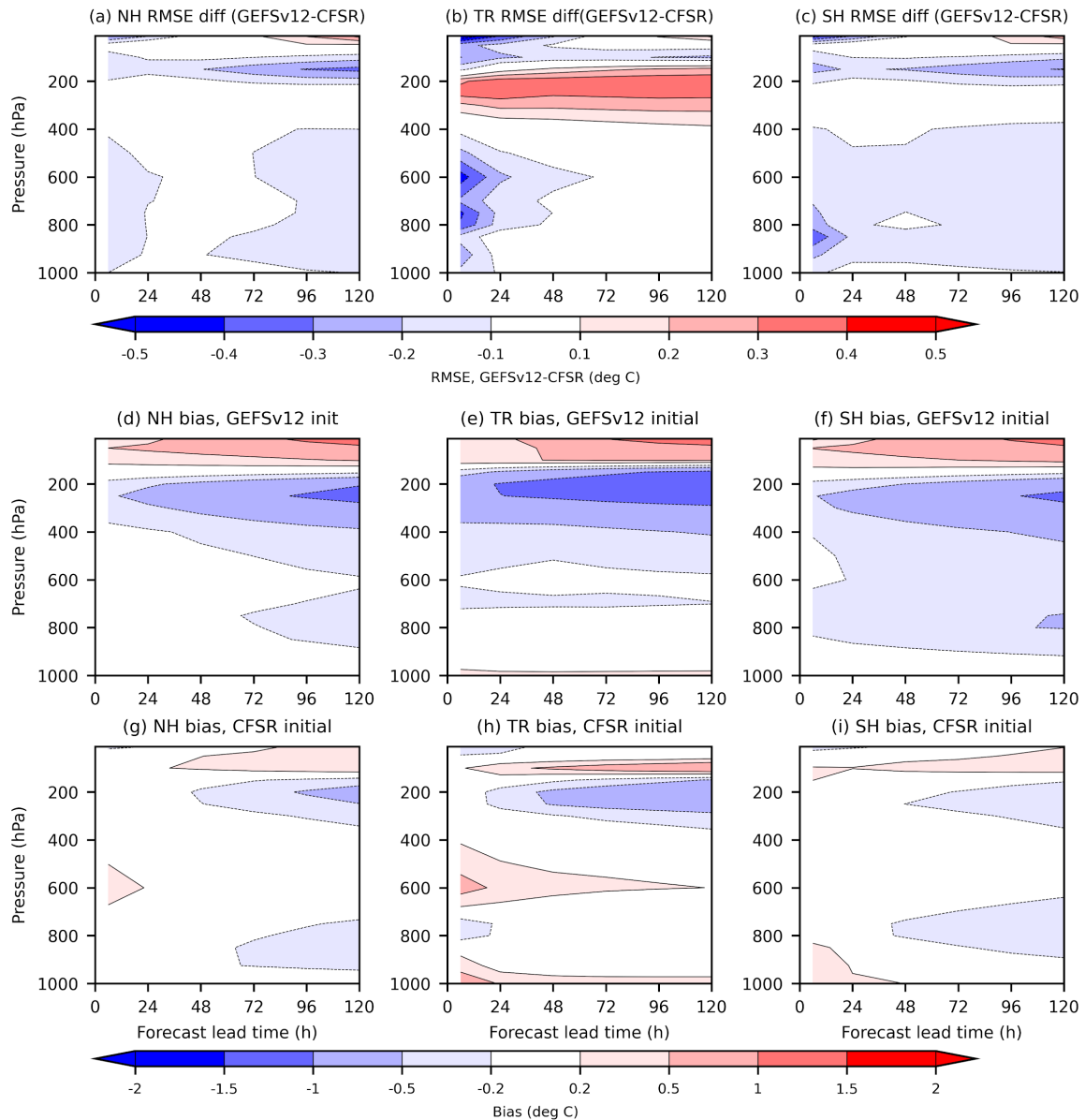


Figure 9. Root-mean-square error (RMSE) and bias characteristics of deterministic FV3-based GFS temperature forecasts initialized from GEFSv12 reanalysis and CFSR initial conditions relative to ERA-Interim reanalyses. Average differences are shown as a function of forecast lead time (x axis) and pressure level (y axis). RMSE differences (GEFSv12 initialized minus CFSR initialized) for the (a) Northern hemisphere (20°- 90°N), (b) tropics (20°S - 20°N), and (c) Southern hemisphere (90°S - 20°S). Bias of GEFSv12 initialized forecasts for the (d) Northern hemisphere, (b) tropics, and (c) Southern hemisphere. Bias of CFSR initialized forecasts for the (g) Northern hemisphere, (h) tropics, and (i) Southern hemisphere.

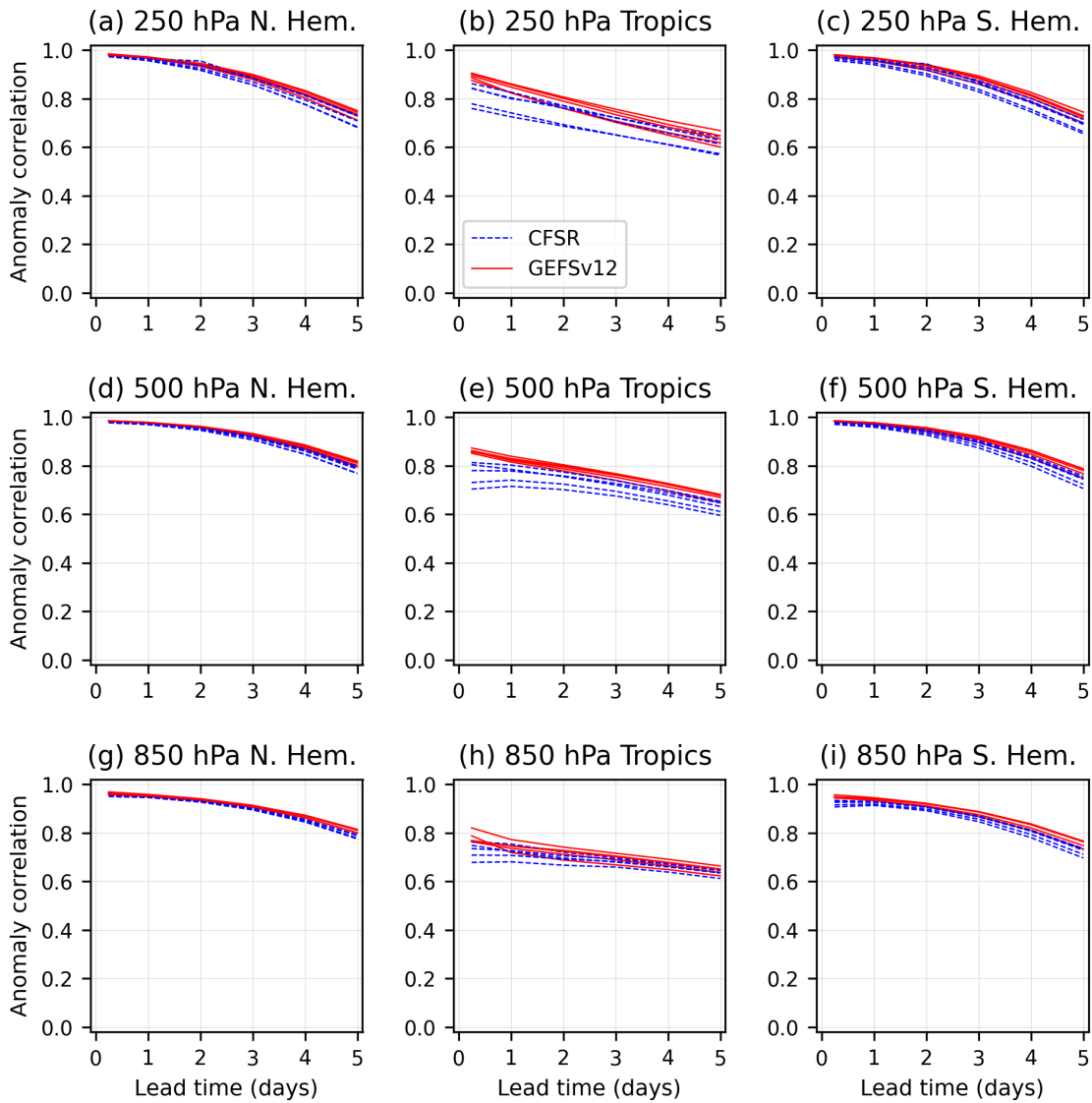
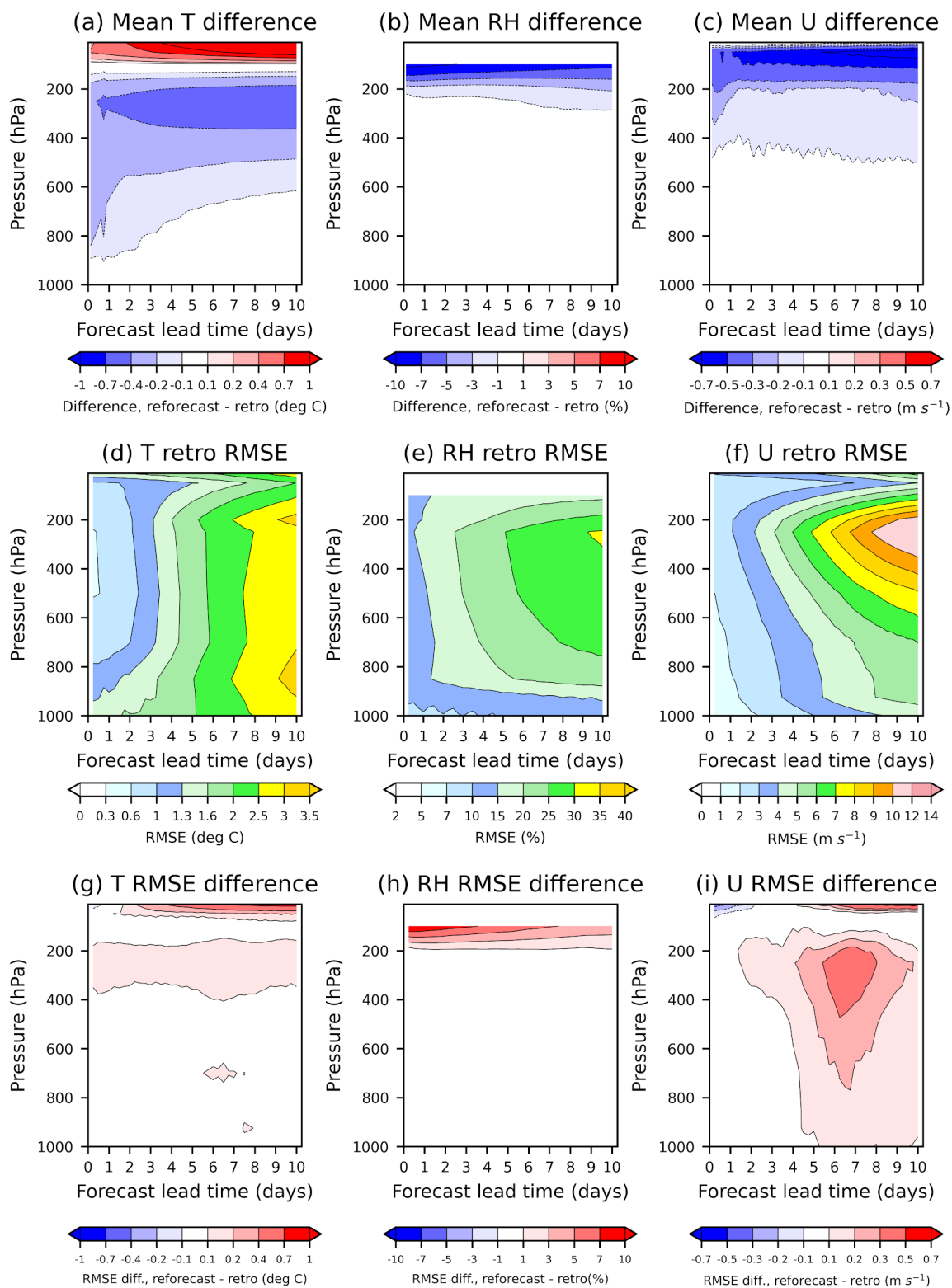


Figure 10: Anomaly correlation (AC) “dieoff” curves of temperature. Different rows represent different vertical levels, 250 hPa, 500 hPa, and 850 hPa, respectively. Different columns represent the northern hemisphere (20°- 90°N), tropics (20°S - 20°N), and southern hemisphere (90°S - 20°S), respectively. Red lines provide AC averages over each stream with GEFSv12 initialization of the FV3 GFS deterministic forecast, neglecting the first (spin-up) year. Blue dashed lines provide averages over each stream with CFSR initialized FV3 GFS deterministic forecast.

GEFSv12 retro and reforecast differences and RMSE



846

847

(caption next page)

Figure 11: Comparisons between mean of 5-member C384 ensemble predictions initialized from the GEFSv12 reanalysis and the “retro” analysis for the period (Dec 2017 to Nov 2019). Data are plotted as a function of forecast lead time (abscissa) and vertical pressure level (ordinate); RH data were not available above 100 hPa. The top row shows the global-mean difference in reanalysis-initialized forecasts minus retro-initialized forecasts for (a) temperature, (b) relative humidity, and (c) u-wind component. The second row shows the retro-run prediction global and ensemble-mean RMSE for (d) temperature, (e), relative humidity, and (f) u-wind component. The bottom row shows difference in ensemble-mean RMSE of predictions from the reforecast minus those from the retro.

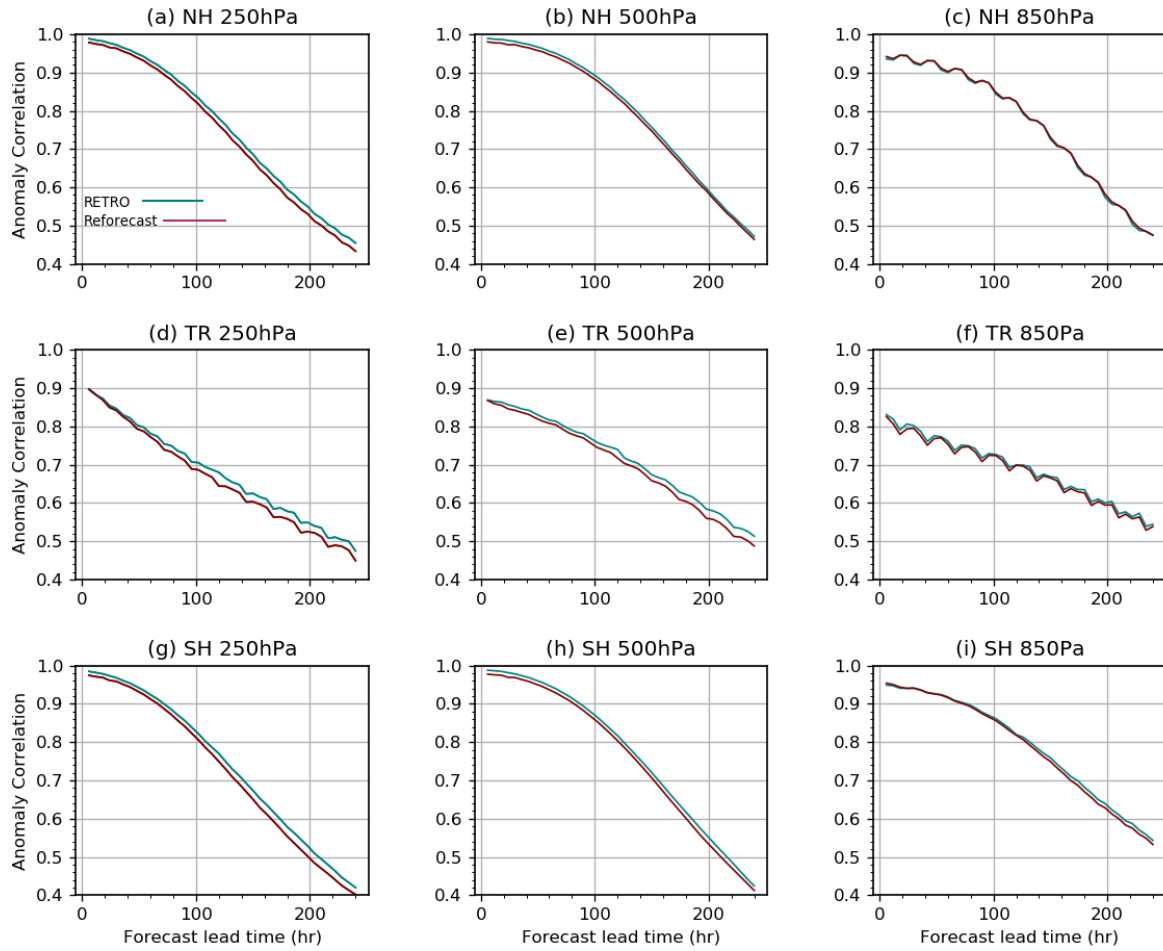


Figure 12: Anomaly correlation dieoff curves for forecast temperature in the Northern Hemisphere (top row), Tropics (middle row), and Southern Hemisphere (bottom row), for 250 hPa (left column), 500 hPa (center column), and 850 hPa (right column). Reforecasts are presented in red and retro simulations in cyan.